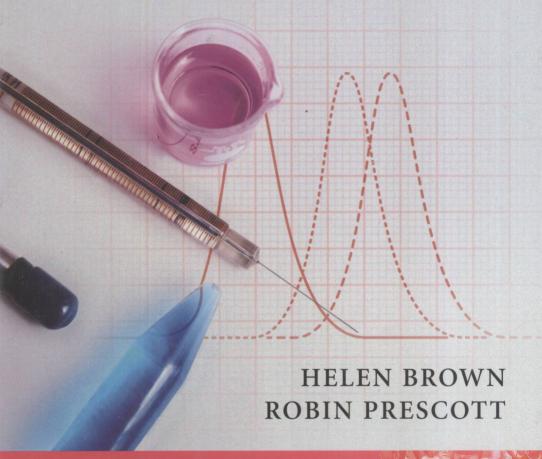
SECOND EDITION

Applied Mixed Models in Medicine





R311 B819 E2

Applied Mixed Models in Medicine

Second Edition

Helen Brown

NHS Scotland Edinburgh UK

and

Robin Prescott

Medical Statistics Unit Public Health Sciences University of Edinburgh UK







John Wiley & Sons, Ltd

Copyright © 2006

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The Publisher is not associated with any product or vendor mentioned in this book.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossev-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data:

Brown, Helen, 1962-

Applied mixed models in medicine / Helen Brown and Robin Prescott.

2nd ed.

p. ; cm.

Includes bibliographical references and index.

ISBN-13:978-0-470-02356-3 (cloth: alk. paper)

ISBN-10:0-470-02356-2 (cloth: alk. paper)

 $\begin{array}{ll} 1.\,Medicine-Research-Statistical\,methods. & 2.\,Medicine-Statistical\,\\ methods. & 3.\,Statistics. & I.\,Prescott,Robin. & II.\,Title. \end{array}$

[DNLM: 1. Medicine. 2. Statistics – methods. 3. Models, Statistical.

WA 950B878a 2006]

R853.S7B76 2006

610.72'7-dc22

2005036661

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13: 978-0-470-02356-3 (HB)

ISBN-10: 0-470-02356-2 (HB)

Typeset in 10/12pt Photina by Laserwords Private Limited, Chennai, India Printed and bound in Great Britain by TJ International, Padstow, Cornwall This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Applied Mixed Models in Medicine

Second Edition

Statistics in Practice

Advisory Editor

Stephen Senn University of Glasgow, UK

Founding Editor

Vic Barnett Nottingham Trent University, UK

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Preface to Second Edition

Analysis of variance and regression analysis have for many years been the mainstay of statistical modelling. These techniques usually have as a basic assumption that the residual or error terms are independently and identically distributed. Mixed models are an important new approach to modelling which allows us to relax the independence assumption and take into account more complicated data structures in a flexible way. Sometimes this interdependence of observations is modelled directly in a mixed model. For example, if a number of repeated measurements are made on a patient, mixed models allow us to specify a pattern for the correlation between these measurements. In other contexts, such as the cross-over clinical trial, specifying that patient effects are normally distributed, rather than fixed as in the classical approach, induces observations on the same patient to be correlated.

There are many benefits to be gained from using mixed models. In some situations the benefit will be an increase in the precision of our estimates. In others, we will be able to make wider inferences. We will sometimes be able to use a more appropriate model which will give us greater insight into what underpins the structure of the data. However, it is only the relatively recent availability of software in versatile packages such as SAS® which has made these techniques widely accessible. It is now important that suitable information on their use becomes available so that they may be applied confidently on a routine basis.

Our intention in this book is to put all types of mixed model into a general framework and to consider the practical implications of their use. We aim to do this at a level that can be understood by applied statisticians and numerate scientists. Greatest emphasis is placed on skills required for the application of mixed models and interpretation of the results. An in-depth understanding of the mathematical theory underlying mixed models is not essential to gain these, but an awareness of the practical consequences of fitting different types of mixed models is necessary. While many publications are available on various aspects of mixed models, these generally relate to specific types of model and often differ in their use of terminology. Such publications are not always readily comprehensible

to the applied statisticians who will be the most frequent users of the methods. An objective of this book is to help overcome this deficit.

Examples given will primarily relate to the medical field. However, the general concepts of mixed models apply equally to many other areas of application, e.g. social sciences, agriculture, official statistics. (In the social sciences mixed models are often referred to as 'multi-level' models.) Data are becoming easier to collect with the consequence that datasets are now often large and complex. We believe that mixed models provide useful tools for modelling the complex structures that occur in such data.

The second edition of this book retains the same structure as the first edition, but there are substantial changes to reflect the advances in software. All of the examples from the first edition have been reworked in Version 9 of SAS using, where relevant, the newly available PROC GLIMMIX for fitting generalised linear mixed models and categorical mixed models. There are also new sections on bioequivalence studies and cluster randomised trials, and a complicated laboratory experiment has been added to the examples. There has been a general updating throughout to reflect recent literature and software changes.

Chapter 1 provides an introduction to the capabilities of mixed models, defines general concepts and gives their basic statistical properties. Chapter 2 defines models and fitting methods for normally distributed data. Chapter 3 first introduces generalised linear models which can be used for the analysis of data which are binomial or Poisson or from any other member of the exponential family of distributions. These methods are then extended to incorporate mixed models concepts under the heading of generalised linear mixed models. The fourth chapter then examines how mixed models can be applied when the variable to be analysed is categorical. The main emphasis in these chapters, and indeed in the whole book, is on classical statistical approaches to inference, based on significance tests and confidence intervals. However, the Bayesian approach is also introduced, since it has several potential advantages and its use is becoming more widespread. Although the overall emphasis of the book is on the application of mixed models techniques, these chapters can also be used as a reference guide to the underlying theory of mixed models.

Chapters 5–7 consider the practical implications of using mixed models for particular designs. Each design illustrates a different feature of mixed models.

Multi-centre trials and meta-analyses are considered in Chapter 5. These are examples of hierarchical data structures and the use of a mixed model allows for any additional variation in treatment effects occurring between centres (or trials) and hence makes results more generalisable. The methods shown can be applied equally to any type of hierarchical data.

In Chapter 6 the uses of covariance pattern models and random coefficients models are described using the repeated measures design. These approaches take into account the correlated nature of the repeated observations and give more appropriate treatment effect estimates and standard errors. The material in this

chapter will apply equally to any situation where repeated observations are made on the same units.

Chapter 7 considers cross-over designs where each patient may receive several treatments. In this design more accurate treatment estimates are often achieved by fitting patient effects as random. This improvement in efficiency can occur for any dataset where a fixed effect is 'crossed' with a random effect.

In Chapter 8 a variety of other designs and data structures are considered. These either incorporate several of the design aspects covered in Chapters 5-7 or have structures that have arisen in a more unplanned fashion. They help to illustrate the broad scope of application of mixed models. This chapter includes the new section on bioequivalence studies and cluster randomised trials.

Chapter 9 gives information on software available for fitting mixed models. Most of the analyses in the book are carried out using PROC MIXED in SAS, supplemented by PROC GENMOD and PROC GLIMMIX, and this chapter introduces the basic syntax for these procedures. This information should be sufficient for fitting most of the analyses described but the full SAS documentation should be referenced for those who wish to use more complex features. The SAS code used for most of the examples is supplied within the text. Additionally, the example datasets and SAS code may be obtained electronically from www.chs.med.ac.uk/phs/mixed.

This book has been written to provide the reader with a thorough understanding of the concepts of mixed models and we trust it will serve well for this purpose. However, readers wishing to take a shortcut to the fitting of normal mixed models should read Chapter 1 for an introduction, Section 2.4 for practical details, and the chapter relevant to their design. To fit non-normal or categorical mixed models, Section 3.3 or Section 4.4 should be read in addition to Section 2.4. In an attempt to make this book easier to use, we have presented at the beginning of the text a summary of the notation we have used, while at the end we list some key definitions in a glossary.

Our writing of this book has been aided in many ways. The first edition evolved from a constantly changing set of course notes that accompanied a three-day course on the subject, run regularly over the previous six years. The second edition has been helped by many individuals who were kind enough to comment on the first edition, including the identification of some errors that had slipped in, and by further participants at our courses who have contributed to discussions and have thereby helped to shape our views. We are also grateful to many other colleagues who have read and commented on various sections of the manuscript and to our colleagues who have allowed us to use their data. We hope that readers will find the resulting book a useful reference in an interesting and expanding area of statistics.

Helen Brown Robin Prescott Edinburgh

Mixed Models Notation

The notation below is provided for quick reference. Models are defined more fully in Sections 2.1, 3.1 and 4.1.

Normal mixed model

```
\begin{aligned} \mathbf{y} &= \mathbf{X}\alpha + \mathbf{Z}\beta + \mathbf{e}, \\ \boldsymbol{\beta} &\sim \mathrm{N}(\mathbf{0}, \mathbf{G}), \\ \mathrm{var}(\mathbf{e}) &= \mathbf{R}, \\ \mathrm{var}(\mathbf{v}) &= \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}. \end{aligned}
```

Generalised linear mixed model

```
\begin{split} & y = \mu + e, \\ & g(\mu) = X\alpha + Z\beta, \\ & \beta \sim \mathrm{N}(0,G), \\ & \mathrm{var}(e) = R, \\ & \mathrm{var}(y) = V = \mathrm{var}(\mu) + R, \\ & \approx BZGZ'B + R \quad \text{(a first-order approximation),} \end{split}
```

where

y = dependent variable,

 $\mathbf{e} = \text{residual error}$.

 $\mathbf{X} = \text{design matrix for fixed effects},$

Z = design matrix for random effects,

 α = fixed effects parameters,

 $\beta = \text{random effects parameters},$

 $\mathbf{R} = \text{residual variance matrix},$

 $\mathbf{G} = \text{matrix of covariance parameters},$

 $\mathbf{V} = \text{var}(\mathbf{y})$ variance matrix,

 μ = expected values,

g = link function,

 \mathbf{B} = diagonal matrix of variance terms (e.g. \mathbf{B} = diag{ $\mu_i(1 - \mu_i)$ } for binary data).

Ordered categorical mixed model

$$\begin{aligned} & y = \mu + e, \\ & \log \mathrm{it}(\mu^{[c]}) = X\alpha + Z\beta, \\ & \beta \sim \mathrm{N}(\mathbf{0}, G), \end{aligned}$$

var(y) is defined as in the GLMM,

where

$$\begin{split} & \boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23}, \dots, \mu_{n1}, \mu_{n2}, \mu_{n3})', \\ & \mu_{ij} = \text{probability observation } i \text{ is in category } j, \\ & \boldsymbol{\mu}^{[c]} = (\mu_{11}^{[c]}, \mu_{12}^{[c]}, \mu_{13}^{[c]}, \mu_{21}^{[c]}, \mu_{22}^{[c]}, \mu_{23}^{[c]}, \dots, \mu_{n1}^{[c]}, \mu_{n2}^{[c]}, \mu_{n3}^{[c]})', \\ & \mu_{ij}^{[c]} = \text{probability } (y_i \leq j) = \sum_{k=1}^{j} \mu_{ik}. \end{split}$$

Statistics in Practice

Human and Biological Sciences

Berger – Selection Bias and Covariate Imbalances in Randomized Clinical Trials

Brown and Prescott – Applied Mixed Models in Medicine, Second Edition

Chevret (Ed) - Statistical Methods for Dose-Finding Experiments

Ellenberg, Fleming and DeMets – Data Monitoring Committees in Clinical Trials: A Practical Perspective

Lawson, Browne and Vidal Rodeiro – Disease Mapping with WinBUGS and MLwiN Lui – Statistical Estimation of Epidemiological Risk

*Marubini and Valsecchi – Analysing Survival Data from Clinical Trials and Observation Studies

Parmigiani – Modeling in Medical Decision Making: A Bayesian Approach

Senn - Cross-over Trials in Clinical Research, Second Edition

Senn – Statistical Issues in Drug Development

Spiegelhalter, Abrams and Myles – Bayesian Approaches to Clinical Trials and Health-Care Evaluation

Whitehead – Design and Analysis of Sequential Clinical Trials, Revised Second Edition

Whitehead – Meta-Analysis of Controlled Clinical Trials

Earth and Environmental Sciences

Buck, Cavanagh and Litton – Bayesian Approach to Interpreting Archaeological Data

Glasbey and Horgan - Image Analysis for the Biological Sciences

Helsel – Nondetects and Data Analysis: Statistics for Censored Environmental Data

McBride – Using Statistical Methods for Water Quality Management Webster and Oliver – Geostatistics for Environmental Scientists

^{*} Now available in paperback

Industry, Commerce and Finance

Aitken – Statistics and the Evaluation of Evidence for Forensic Scientists, Second Edition

Balding - Weight-of-evidence for Forensic DNA Profiles

Lehtonen and Pahkinen – Practical Methods for Design and Analysis of Complex Surveys, Second Edition

Ohser and Mücklich – Statistical Analysis of Microstructures in Materials Science Taroni, Aitken, Garbolino and Biedermann – Bayesian Networks and Probabilistic Inference in Forensic Science

Contents

P	Preface to Second Edition			xiii
Mixed Model Notations			xvii	
1	Intr	oductio	on	1
	1.1	The Us	se of Mixed Models	1
	1.2	Introd	uctory Example	3
		1.2.1	Simple model to assess the effects of treatment	
			(Model A)	3
		1.2.2	A model taking patient effects into account	
			(Model B)	6
		1.2.3	Random effects model (Model C)	6
		1.2.4	Estimation (or prediction) of random effects	11
	1.3		ti-Centre Hypertension Trial	12
		1.3.1	Modelling the data	13
		1.3.2	Including a baseline covariate (Model B)	13
		1.3.3	Modelling centre effects (Model C)	16
		1.3.4	Including centre-by-treatment interaction effects	
			(Model D)	16
		1.3.5	Modelling centre and centre treatment effects as	
			random (Model E)	17
	1.4	_	ted Measures Data	18
		1.4.1	Covariance pattern models	19
		1.4.2	Random coefficients models	20
	1.5		about Mixed Models	22
		1.5.1	What is a mixed model?	22
		1.5.2	Why use mixed models?	23
		1.5.3	Communicating results	24
		1.5.4	Mixed models in medicine	25
		1.5.5	Mixed models in perspective	25
	1.6		Useful Definitions	27
		1.6.1	Containment	27

VI	Contents

		1.6.2	Balance	28
		1.6.3	Error strata	30
2	Norn	nal Mix	ed Models	33
_	2.1		Definition	33
	2.1	2.1.1	The fixed effects model	34
		2.1.2	The mixed model	36
		2.1.3	The random effects model covariance structure	38
		2.1.4	The random coefficients model covariance structure	41
		2.1.5	The covariance pattern model covariance structure	43
	2.2		Fitting Methods	45
		2.2.1	The likelihood function and approaches to its	
			maximisation	45
		2.2.2	Estimation of fixed effects	49
		2.2.3	Estimation (or prediction) of random effects and	
			coefficients	50
		2.2.4	Estimation of variance parameters	52
	2.3	The Ba	yesian Approach	56
		2.3.1	Introduction	57
		2.3.2	Determining the posterior density	58
		2.3.3	Parameter estimation, probability intervals and	
			<i>p</i> -values	59
		2.3.4	Specifying non-informative prior distributions	61
		2.3.5	Evaluating the posterior distribution	66
	2.4	Practic	cal Application and Interpretation	70
		2.4.1	Negative variance components	70
		2.4.2	Accuracy of variance parameters	74
		2.4.3	Bias in fixed and random effects standard errors	75
		2.4.4	Significance testing	76
		2.4.5	Confidence intervals	79
		2.4.6	Model checking	79
		2.4.7	Missing data	81
	2.5	Examp		83
		2.5.1	Analysis models	83
		2.5.2	Results	85
		2.5.3	Discussion of points from Section 2.4	87
3	Gene	eralised	l Linear Mixed Models	107
	3.1	Genera	alised Linear Models	108
		3.1.1	Introduction	108
		3.1.2	Distributions	109
		3.1.3	The general form for exponential distributions	111
		3.1.4	The GLM definition	112
		3.1.5	Fitting the GLM	115

			Contents	vii
		3.1.6	Expressing individual distributions in the general	
		3.1.0	exponential form	117
		3.1.7	Conditional logistic regression	119
	3.2		lised Linear Mixed Models	120
	J.2	3.2.1	The GLMM definition	120
		3.2.2	The likelihood and quasi-likelihood functions	121
		3.2.3	Fitting the GLMM	124
	3.3		al Application and Interpretation	128
		3.3.1	Specifying binary data	128
		3.3.2	Uniform effects categories	129
		3.3.3	Negative variance components	130
		3.3.4	Fixed and random effects estimates	130
		3.3.5	Accuracy of variance parameters and random	
			effects shrinkage	131
		3.3.6	Bias in fixed and random effects standard errors	132
		3.3.7	The dispersion parameter	133
		3.3.8	Significance testing	135
		3.3.9	Confidence intervals	136
		3.3.10	Model checking	136
	3.4	Examp		137
		$3.4.1^{-1}$	Introduction and models fitted	137
		3.4.2	Results	139
		3.4.3	Discussion of points from Section 3.3	142
4	Mixe	ed Mode	els for Categorical Data	153
_	4.1		l Logistic Regression (Fixed Effects Model)	153
	4.2		Ordinal Logistic Regression	158
	1.4	4.2.1	Definition of the mixed ordinal logistic regression	
		1.2.1	model	158
		4.2.2	Residual variance matrix	159
		4.2.3		161
		4.2.4	Likelihood and quasi-likelihood functions	162
		4.2.5	Model fitting methods	162
	4.3		Models for Unordered Categorical Data	163
	2.0	4.3.1	The \mathbf{G} matrix	165
		4.3.2	The R matrix	166
		4.3.3	Fitting the model	166
	4.4		cal Application and Interpretation	166
	577/53/(F)	4.4.1	Expressing fixed and random effects results	167
		4.4.2	The proportional odds assumption	167
		4.4.3	Number of covariance parameters	167
		4.4.4	Choosing a covariance pattern	168
		4.4.5	Interpreting covariance parameters	168
		446	Checking model assumptions	168

viii	Contents

		4.4.7	The dispersion parameter	168
		4.4.8	Other points	168
	4.5	Examp	le	169
5	Mult	i-Centr	re Trials and Meta-Analyses	183
	5.1	Introdu	uction to Multi-Centre Trials	183
		5.1.1	What is a multi-centre trial?	183
		5.1.2	Why use mixed models to analyse multi-centre	
			data?	184
	5.2		aplications of using Different Analysis Models	184
		5.2.1	Centre and centre-treatment effects fixed	184
		5.2.2	,	185
		5.2.3	Centre and centre-treatment effects random	186
		5.2.4	Centre effects random, centre-treatment effects	
	5 2		omitted	187
	5.3	_	le: A Multi-Centre Trial	188
	5.4		cal Application and Interpretation	195
		5.4.1	200	195
		5.4.2	Generalisation	195
		5.4.3	Number of centres	196
		5.4.4	Centre size	196
		5.4.5	Negative variance components	196
		5.4.6	Balance	197
	5.5		e Size Estimation	197
			Normal data	197
	- (5.5.2	Non-normal data	201
	5.6		Analysis	203
	5.7		le: Meta-analysis	204
			Analyses	204
		5.7.2	Results	205
		5.7.3	Treatment estimates in individual trials	206
6	Repe	ated M	easures Data	215
	6.1	Introdu	action	215
		6.1.1	SACTOR CONTRACTOR CONT	215
		6.1.2	Analysis objectives	216
		~ . – . –	Fixed effects approaches	216
		6.1.4	Mixed models approaches	217
	6.2	Covari	ance Pattern Models	218
		6.2.1	Covariance patterns	219
		6.2.2	Choice of covariance pattern	223
		6.2.3	Choice of fixed effects	225
		6.2.4	General points	226
	6.3	Examp	le: Covariance Pattern Models for Normal Data	228

		6.3.1	Analysis models	229
		6.3.2	Selection of covariance pattern	229
		6.3.3	Assessing fixed effects	231
		6.3.4	Model checking	231
	6.4	Example	: Covariance Pattern Models for Count Data	237
		6.4.1	Analysis models	238
		6.4.2	Analysis using a categorical mixed model	242
	6.5	Random	Coefficients Models	245
		6.5.1	Introduction	245
		6.5.2	General points	247
		6.5.3	Comparisons with fixed effects approaches	249
	6.6	Example	es of Random Coefficients Models	249
		6.6.1	A linear random coefficients model	250
		6.6.2	A polynomial random coefficients model	252
	6.7	Sample	Size Estimation	267
			Normal data	267
		6.7.2	Non-normal data	269
		6.7.3	Categorical data	270
7	Cros	s-Over T		271
	7.1	Introdu		271
	7.2		ages of Mixed Models in Cross-Over Trials	272
	7.3		/BA Cross-Over Trial	272
			Example: AB/BA cross-over design	275
	7.4	_	Order Complete Block Designs	279
		7.4.1		279
		7.4.2	Example: four-period, four-treatment cross-over	
			trial	279
	7.5	Incomp	olete Block Designs	284
		7.5.1	The three-treatment, two-period design (Koch's	201
			design)	284
		7.5.2	Example: two-period cross-over trial	285
	7.6	Optima	l Designs	287
		7.6.1	Example: Balaam's design	287
	7.7		ance Pattern Models	290
		7.7.1	Structured by period	290
		7.7.2	Structured by treatment	290
		7.7.3	Example: four-way cross-over trial	291
	7.8		is of Binary Data	299
	7.9		is of Categorical Data	303
	7.10	Use of I	Results from Random Effects Models in Trial Design	307
		7.10.1	Example	308
	7.11	Genera	ll Points	308

Contents

ix